



**University of
Zurich**^{UZH}

**Zurich Open Repository and
Archive**

University of Zurich
University Library
Strickhofstrasse 39
CH-8057 Zurich
www.zora.uzh.ch

Year: 2017

Newspaper coverage of female candidates during election cam paigns: evidence from a structural topic model

Gilardi, Fabrizio ; Wüest, Bruno

Abstract: This paper studies how female candidates are covered by newspapers during election campaigns. Previous studies have generally found that the media portray men and women quite similarly, but they have tended to use relatively selective sources and to focus on the amount of coverage and its tone, but not on its content. We aim to gain a more nuanced understanding of this phenomenon by focusing on the Swiss national elections 2015, relying on an almost comprehensive sample of news items covering most of the duration of the campaign. Our dataset includes about 205,000 documents from 70 sources, covering over 3,900 male and female politicians. We analyze these texts first with a focus on the amount of coverage and second using structural topic models, a natural language processing technique that allows us to identify inductively the themes (topics) of newspaper coverage and, importantly, how the candidates' gender is linked to both the topics and the language used to discuss them. Results reveal a gender bias in media attention only for non-incumbent candidates. Moreover, some topics, such as the one referring to the election campaign, is discussed more frequently in connection with female candidates, while others, such as topics focusing on personality and political profiles, are more prevalent in articles mentioning men. We uncover no major gender bias in the language used for these topics.

Posted at the Zurich Open Repository and Archive, University of Zurich

ZORA URL: <https://doi.org/10.5167/uzh-143863>

Conference or Workshop Item

Draft Version

Originally published at:

Gilardi, Fabrizio; Wüest, Bruno (2017). Newspaper coverage of female candidates during election cam paigns: evidence from a structural topic model. In: European Conference on Politics and Gender. Annual meeting, Lausanne, 8 June 2017 - 10 June 2017, 1-22.

Newspaper coverage of female candidates during election campaigns: Evidence from a structural topic model*

Fabrizio Gilardi[†]

Bruno Wueest[‡]

First draft

June 7, 2017

Abstract

This paper studies how female candidates are covered by newspapers during election campaigns. Previous studies have generally found that the media portray men and women quite similarly, but they have tended to use relatively selective sources and to focus on the amount of coverage and its tone, but not on its content. We aim to gain a more nuanced understanding of this phenomenon by focusing on the Swiss national elections 2015, relying on an almost comprehensive sample of news items covering most of the duration of the campaign. Our dataset includes about 205,000 documents from 70 sources, covering over 3,900 male and female politicians. We analyze these texts first with a focus on the amount of coverage and second using structural topic models, a natural language processing technique that allows us to identify inductively the themes (topics) of newspaper coverage and, importantly, how the candidates' gender is linked to both the topics and the language used to discuss them. Results reveal a gender bias in media attention only for non-incumbent candidates. Moreover, some topics, such as the one referring to the election campaign, is discussed more frequently in connection with female candidates, while others, such as topics focusing on personality and political profiles, are more prevalent in articles mentioning men. We uncover no major gender bias in the language used for these topics.

*We thank Sarah Bütikofer, Adrian van der Lek and Fionn Gantenbein for excellent research assistance. The financial support of the Swiss National Science Foundation (grant nr. 10FI14_150999) is gratefully acknowledged.

[†]Department of Political Science, University of Zurich (<http://www.fabriziogilardi.org/>).

[‡]Department of Political Science, University of Zurich (<http://www.bruno-wueest.ch/>).

1 Introduction

There are many ways in which gender bias affects elections, including differential political ambition for men and women, biases in the recruiting strategies of parties, and voter behavior. We focus on the way in which the media report on male and female candidates. How female candidates are portrayed in the media is arguably of considerable importance both for their electoral prospects and for their decision to run for office in the first place. The literature has shown that media bias has decreased in the past decades, and scholars disagree about the nature and extent of remaining gender differences.

This study aims to gain a nuanced understanding of gender bias in media coverage of elections by relying on an almost comprehensive dataset of news items covering the 2015 Swiss national elections, including over 200,000 documents from 70 sources, matched with information on all 3,931 candidates. Our analysis proceeds in three steps.

First, we focus on media attention. The unit of analysis in this part is the candidate and the dependent variable the number of times he or she was mentioned in newspaper articles during the campaign, as well as the number of documents including at least one mention. We find a gender bias only for non-incumbents. Male and female incumbents received essentially the same attention from the media.

Second, we use structural topic models to identify the topics of media coverage and, especially, how the prevalence of each topic correlates with the gender of the candidates mentioned in the texts. We find that the election campaign was discussed more frequently in texts mentioning only men than in texts mentioning only women. By contrast, personality and political profiles are discussed more frequently in connection with men than with women.

Third, always relying on structural topic models, we look at the language used to discuss each topic and how it co-varies with the gender of the candidates mentioned in the texts. We find few gender differences here.

We emphasize that this paper reports the output of preliminary analyses. Because the results are bound to change at least to some extent, the focus is more on the nature of the results that we can offer than on their details.

2 Previous Studies

There is a relatively large literature on how female candidates are portrayed in the media, relative to men. Studies have focused on two main aspects: first, attention (e.g., the number of stories mentioning a candidate) and second, content (how a candidate is portrayed). When looking at content, studies typically differentiate among several categories, such as issues, horse race, and candidate traits. The consensus seems to be that biases in coverage have sharply decreased over time, but there is disagreement over the remaining extent of bias as well as its nature and causes.

Early studies noted a significant bias in coverage against women (Kahn and Goldenberg, 1991; Kahn, 1994). However, a decrease in bias was noted as early as the 1990s: “Analyzing statewide campaigns in 1994, I find much smaller coverage differences than in studies relying on pre-1990 data” (Smith, 1997, 71). Later studies find that the bias keep getting smaller, while uncovering more subtle ways in which coverage disadvantages women (Bystrom, Robertson and Banwart, 2001; Banwart, Bystrom and Robertson, 2003; Heldman, Carroll and Olson, 2005; Jalalzai, 2006; Fowler and Lawless, 2009; Gershon, 2012; Hayes, Lawless and Baitinger, 2014; Baitinger, 2015; Escobar-Lemmon et al., 2016; Ward, 2016).

Kittilson and Fridkin (2008) is representative of the literature in several ways.¹ First, it relies on a small sample (354 articles) that were coded manually. Second, it finds little bias on several important dimensions: it “any significant gender differences in 1) the amount and prominence of candidate coverage, 2) the amount of attention given to the candidates’ viability, 3) the focus on the candidates’ family background, and 4) the tone of coverage” (Kittilson and Fridkin, 2008, 381). Third, it notes remaining imbalances in reporting, notably in terms of the gender stereotypes used by journalists.

Dunaway et al. (2013) is one of the largest studies, including over 10,000 news stories. It focuses on the content of coverage in election with female candidates, compared with all-male races, and finds that the former tend to be more focused on traits rather than issues.

Lühiste and Banducci (2016) is another large study, relying on the 2009 European Election Study’s Media Content Data, with candidate level data on media coverage from 25 European Union member states. It finds a small but persistent gender gap in the amount of coverage, robust to controlling for candidate viability.

Hayes and Lawless (2015) focuses on the US, with data on 4,748 news stories in 342 districts, for a total of 108 female and 555 male candidates. It finds almost no gender gap in coverage, neither for

¹Unlike most studies, it compares three countries (Australia, Canada, United States).

frequency nor for the traits emphasized in the stories.

Like our study, Fiechtner, Schönhagen and Puppis (2016) analyzes the gender gap in the coverage of the 2015 Swiss national elections (both newspapers and radio/TV). Like other studies, it relies on a small sample (580 texts) that are coded manually. The study finds that female candidates are underrepresented, relative to their presence on electoral lists. However, the study finds no differences in the ways female and male candidates are portrayed.

3 Methodology

3.1 Corpus

Our analysis of the newspaper coverage of female candidates is based on the Selects Media Analyses 2015 (Wüest et al., 2016). For the moment, we are only able to process and analyze the 70 german-language newspapers in the corpus (see appendix A1 for a list and key figures of the corpus).² National election campaigns are usually started with keynote speeches of candidates at the festivities surrounding the Swiss national holiday (first of August). This is why the original corpus of 205,607 newspaper articles was collected from the national holiday until election day, which was October 18 in 2015. Thanks to a direct access to the repositories of the Swiss Media Service (<https://smd.ch/SMDView/>), we could retrieve virtually all documents published by the respective news sources during this election campaign. The corpus comprises a large variety of sources: tabloid magazines such as *Schweizer Familie*, nationally important newspapers such as the *Neue Zürcher Zeitung* (quality press) or the *Blick* (tabloid), regional and local newspapers such as the *Zofinger Tagblatt*, online news sources such as *20 Minuten online* to specialized news outlets such as the *Finanz und Wirtschaft* (business news). Hence, we are confident to have a comprehensive enough data basis to map the general newspaper coverage of candidates in the German-speaking parts of Switzerland.

We use print media rather than television or radio programs partly for technical reasons but especially because they generally report more extensively on political matters than do on-air media (Druckman, 2005, 469). And despite strong convergence pressures, the newspaper market in Switzerland still is comparatively fragmented. Moreover, national elections in Switzerland essentially are cantonal elections, in which cantonal party sections have a decisive say in the selection of candidates and a commanding lead over elections campaigns in their constituency. Especially when it comes to cantonal political news

²In the future, we plan to extend our empirical analysis to the french-language newspapers.

such as the election campaign, we can therefore assume that newspapers convey the richest information.

To identify the relevant documents we relied on dictionary lookups and regular expressions. Concretely, we selected all documents featuring at least one name in a dictionary compiled from a) the official list of candidates in the federal election 2015 running for seats in the National Council or the Council of States, b) a custom-compiled list of Federal Councillors, party presidents and resigning incumbents. The final dictionary features 3,913 politicians. We only searched for the combination of first and last name³, which greatly reduces the risk of false positives. Such a filter strategy is especially feasible since all entities of interest – i.e. the names of politicians – are readily available and, with only a few exceptions, precisely searchable. From the originally 205,607 newspaper articles, about 10% contained at least one candidate. We therefore end up with a corpus of 20,104 texts.

3.2 Media Attention

For the first analysis on media attention to female and male candidates, we estimate negative binomial regressions at the individual level, that is, at the level of candidates. Here, the unit of analysis is the candidate and we use two dependent variables: first, the number of mentions for a given candidate, and second, the number of documents in which a candidate is mentioned. We estimate negative binomial models because these two variables are highly over-dispersed. Some few persons like the Federal Councillors have a lot of mentions, while many candidates have few or no mentions at all⁴. We consider the following explanatory variables: the gender of a candidate, whether he or she is an incumbent, the interaction between gender and incumbency, whether the person mentioned is a Federal Councillor, the party of the candidate, and the canton in which a candidate is running.

3.3 Structural Topic Model

We identify newspaper coverage themes inductively with a structural topic model (STM) (Roberts et al., 2014; Roberts, Stewart and Airola, 2016), which builds on well-established generative topic models, namely the Correlated Topic Model (CTM) (Blei and Lafferty, 2007). Both the CTM and the STM are a mixed-membership models, meaning that they assume that each document consists of a mixture of topics (Grimmer and Stewart, 2013, 283–285). A consequence of the logistic-normal distribution

³The regular expression used was `\\ <[firstname].0,12[lastname]s?\\ >`, which translates to a search for first and last name pairs that have a white space at the beginning and the end, at most twelve characters in between (e.g. a second first name) and an optional genitive s (e.g. ‘Doris Leuthards’) at the end.

⁴948 of the 3913 persons in the analysis are never mentioned, and another 553 are mentioned only once.

underlying these models is that topic prevalences always add up to 1 for each document. Therefore, if a topic has a higher-than-average prevalence in a document, it lowers the prevalence of the other topics.

Concretely, the STM is a hierarchical model in which a document d 's prevalence of each topic is drawn from a logistic-normal linear distribution whose mean is a function of document covariates X_d (Roberts, Stewart and Airolidi, 2016, 991):

$$\theta_d \sim \text{LogisticNormal}(\mu = X_d\gamma, \Sigma).$$

Then, given these document-specific word distributions over the topics, the words of each document ($w_{1...n,d}$) are assigned to the topics:

$$z_{d,n} \sim \text{Multinomial}(\theta_d).$$

Finally, the observed words are drawn based on this topic assignment:

$$w_{d,n} \sim \text{Multinomial}(\beta_{d,z}).$$

where $\beta_{d,z}$ signifies the document-specific distribution over the words w that represent each topic k . It is sampled proportional to the general word distribution m , the deviations of the distributions across topics D_k and covariates D_c as well as the interaction between the two deviations:

$$\beta_{d,w} \propto \exp(m + D_k + D_c + D_k \times D_c).$$

The STM's major innovation is that the prior distribution of topics (prevalence) and words (contents) can be influenced by covariates (Roberts et al., 2014; Roberts, Stewart and Airolidi, 2016). In the following analysis, we correlate the candidates gender with both the topic's prevalence and the word's content.

3.4 Covariates and Estimation

The most important covariate in our analysis measures the *gendered newspaper coverage in the election campaign*. The variable has three items indicating whether only female candidates, only male candidates, or both are mentioned.

The analysis includes several other covariates, namely: (1) a daily trend variable with a B-spline of order 10; (2) newspaper IDs; (3) the log of the number of politicians mentioned in the article; whether at

least one (4) incumbent or (5) federal councillor is mentioned; (6) the party affiliation of the candidates mentioned; and (8) the SMA classification whether an article covers Swiss politics.

We estimate our topic models using the `stm` package in R (Roberts, Stewart and Tingley, 2014). We initialize the models with the spectral algorithm, which is robust to changes in several CTM parameters and starting values (Roberts, Stewart and Airolidi, 2016). To select the number of topics, we evaluated the semantic coherence of the topics using `word2vec` O’Callaghan et al. (2015).⁵ We evaluated 26 models (varying the number of topics from 4 to 30) and found that models with relatively few topics (7 to 13) performed better (see the analysis in Appendix A2). After a qualitative evaluation of the most probable words and documents of the models’ topics in this range, we selected the 12-topic model as the most useful for our analysis at this stage.

4 Results

4.1 Media attention

Table 1 shows that male candidates received more media attention than female candidates, in terms of both the number of mentions and the number of documents in which they were mentioned. Comparing coefficient sizes, the difference between men and women is about 16% of that between incumbents and non-incumbents. However, Table 1 also shows that there is an interaction between gender and incumbency, whereby the gender gap is almost erased when a female candidate is an incumbent. Figure 1 shows the results more clearly, and it becomes apparent that there is no gender gap in the media coverage of incumbents, whereas the gap is substantial for non-incumbents.

4.2 Topics

Figure 2 shows the top-50 words associated with each of our twelve topics, along with labels that we assigned to each topic based on the top-50 words. The order of the figures from left to right and top to bottom reflect the importance of the topics in terms of their average prevalence. The interpretation of the topics is mostly straightforward, which is not self-evident considering that the topics were produced purely inductively, without human input apart from the selection of the covariates and the number of topics.

⁵In contrast to O’Callaghan et al. (2015), we consider not only coherence (the similarity of all word pairs in the same topic) but also discrimination (the inverse similarity of all word pairs across topics) in our evaluation.

	Documents	Documents	Mentions	Mentions
(Intercept)	1.22*** (0.12)	1.22*** (0.12)	1.36*** (0.12)	1.37*** (0.12)
Woman	−0.45*** (0.05)	−0.48*** (0.05)	−0.45*** (0.05)	−0.48*** (0.05)
Incumbent	2.86*** (0.09)	2.73*** (0.11)	2.85*** (0.10)	2.72*** (0.12)
Woman × Incumbent		0.39* (0.20)		0.41** (0.21)
Federal Councillor	5.10*** (0.51)	5.11*** (0.51)	5.10*** (0.54)	5.11*** (0.54)
BDP	0.18 (0.13)	0.18 (0.13)	0.13 (0.14)	0.14 (0.14)
CVP	0.71*** (0.11)	0.71*** (0.11)	0.74*** (0.12)	0.74*** (0.12)
EVP	0.27* (0.15)	0.27* (0.15)	0.29* (0.16)	0.28* (0.16)
FDP	0.86*** (0.11)	0.86*** (0.11)	0.89*** (0.11)	0.89*** (0.11)
GLP	0.21* (0.12)	0.21* (0.12)	0.19 (0.12)	0.18 (0.12)
GPS	0.65*** (0.11)	0.65*** (0.11)	0.64*** (0.12)	0.64*** (0.12)
RL	0.24 (0.15)	0.24 (0.15)	0.24 (0.15)	0.24 (0.15)
RR	−0.30** (0.12)	−0.30*** (0.12)	−0.33*** (0.12)	−0.34*** (0.12)
SP	0.82*** (0.11)	0.81*** (0.11)	0.84*** (0.11)	0.84*** (0.11)
SVP	1.27*** (0.11)	1.27*** (0.11)	1.35*** (0.12)	1.36*** (0.12)
Canton fixed effects	Yes	Yes	Yes	Yes
BIC	23517.95	23522.29	24326.31	24330.56
Log Likelihood	-11593.54	-11591.57	-11997.71	-11995.70
Deviance	4321.46	4321.39	4337.00	4336.84
Num. obs.	3913	3913	3913	3913

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 1: *The influence of gender, incumbency and government affiliation on media attention: log odds ratios and unstandardized coefficients from negative binomial models.*

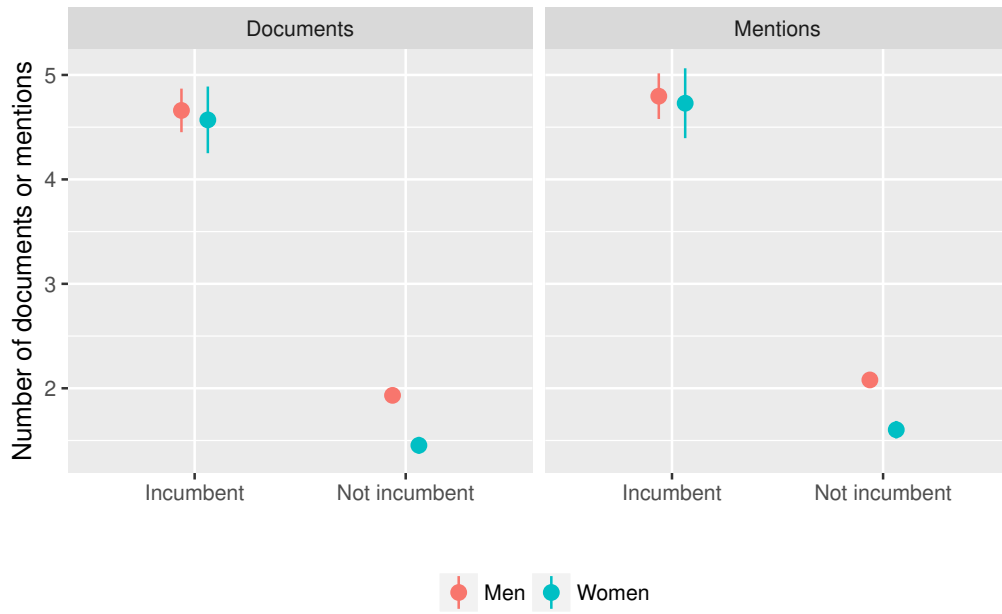


Figure 1: *Media attention to female and male candidates: Fitted values and 95% confidence intervals from regressions shown in table 1.*

We group the twelve topics in two categories, policy and valence. While it is important to explore how candidate are portrayed in terms of competing policy positions, non-policy related factors clearly have an impact on newspaper coverage as well. Therefore, not only substantive coverage, but also coverage related to the personality or the political contest itself are important assets for politicians engaged in electoral campaigns. Such factors include the charisma of the person, the reputation of actors in terms of their competence and trustworthiness, and their media performance (Norris, 2000). Following Laver and Sergenti (Laver and Sergenti, 2011, 164), the term valence will be used here as a catch-all concept summarizing these aspects. Valence attributes generally have an important news value, and mass media are thus keen to exploit these attributes in their competition for audience (Strömbäck and Kaid, 2008).

First, we have seven topics referring to particular policy areas. It is quite clear that these topics cover the substantial aspects of the electoral contest. In the top-50 words, both important actors—such as the World Trade Organisation (“Welhandelsorganisation”) for the trade topic—as well as the most controversial issues and regulations—such as the Schengen-Dublin agreement for the immigration topic—are listed.

The topic *Energy* is centered around the Energy Strategy 2050, a major overhaul of the federal energy

policies debated in parliament during the election 2015. Main aspects of this topic are the promotion of energy efficiency, the emphasis of renewable forms of energies and the nuclear phaseout. The topic *Immigration* is defined by words describing migration and fugitives as well as immigration and asylum policies. *Local development* mainly consists of words depicting spatial planning and traffic regulation. Further, the topic *Media and communication* comprises words related to media enterprises (e.g. “ringi”, which is the stemmed form of the publishing house “Ringier”) and possible conflicts related to political communication (advertising regulation, financing of public media etc.). *Mobility and fiscal policy* is a somewhat mixed topic. It includes words related to traffic and budgetary matters. *Law and order* is again a more clearly defined topic with a lot of references to police and courts. *Trade*, finally, is concerned with policies regulating international economic competition.

The second set of topics focuses on valence, which denote aspects of the candidates’ newspaper coverage related to their personality or the political contest. *Election campaign*, the overall most important topic, describes candidates in terms of possible seat gains or losses as well as electoral alliances. Then, there are two quite similar topics related to personal descriptions of persons. The first one, *personality*, essentially revolves around words describing personal details such as the favorite song or joke. The second one, *Political profiles*, is characterized by terms referring to the electoral dossier or campaign activities such as interviews or pre-election parties. The next topic, *Traditions and folklore*, clearly is related to the start of the election campaign on the national holiday, when candidates are expected to speak at the traditional festivities. Finally, there is the quite surprising topic *Valais* that contains almost exclusively person and geographical names from this particular canton. We have to look more closely into the data in further versions of the analysis why this topic arises. In inductive analyses, such unexpected topics can arise, and, in any case, this is the least important topic in terms of prevalence.

Figure 3 shows how the twelve topics evolved over time. Noticeable is especially the peak of *Election campaign* right before the election, as well as a bump in *Political profiles* from mid-September until mid-October. *Traditions and folklore* peaks at the beginning of August, linked with the Swiss national holiday on the 1st of August. *Immigration* has a mild but noticeable downwards trend. *Energy* is punctuated, probably in connection with discussion in parliament.

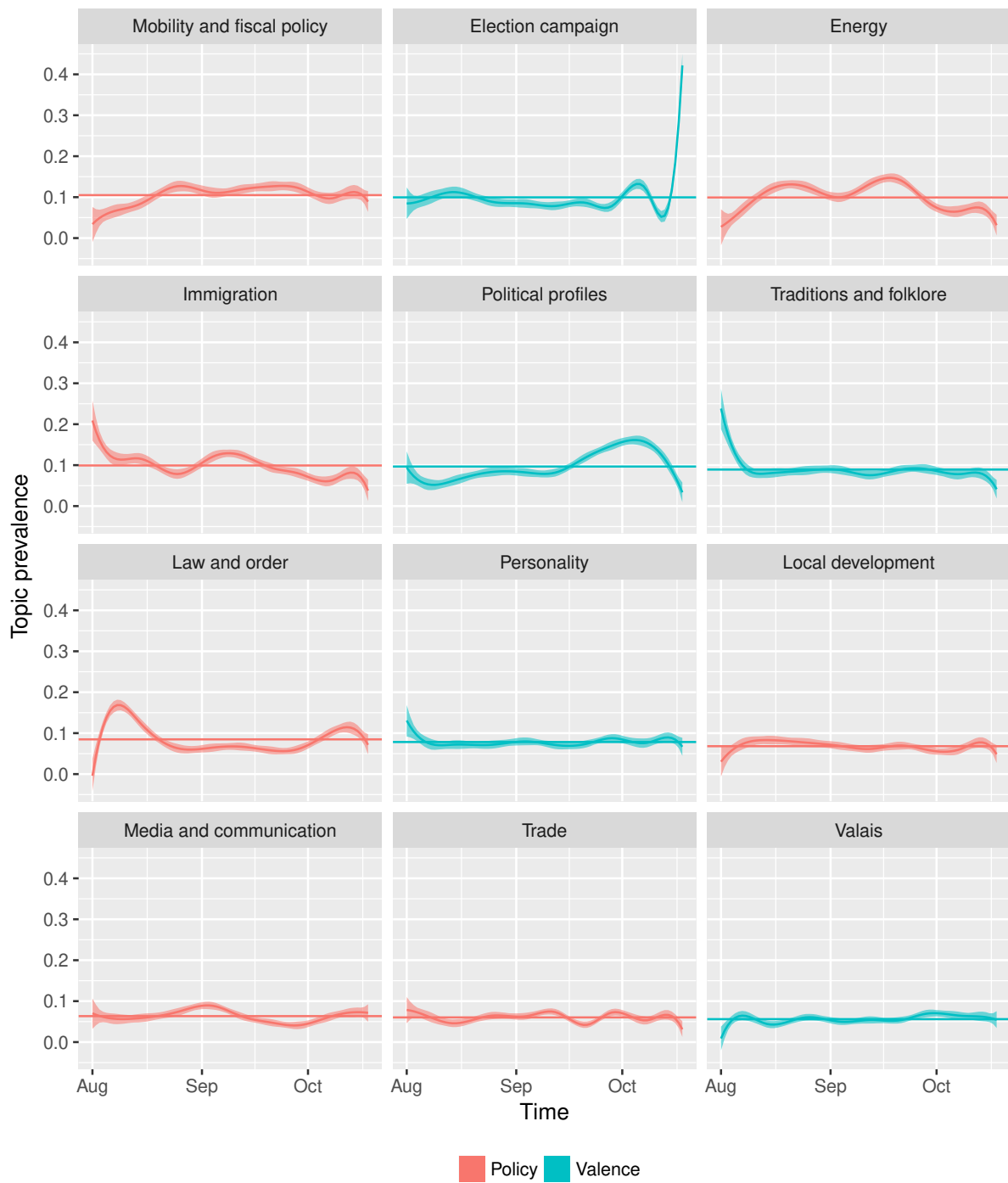


Figure 3: *Topic prevalence over time. Topics are sorted by decreasing average prevalence. Horizontal lines show average prevalence for each topic over the observation period.*

4.3 Gender

Figure 4 shows correlations between the prevalence of each topic with the gender of the candidates mentioned in the articles—specifically, whether an article mentions only women, only men, or both men and women. For instance, the topic *Election campaign* is most prevalent when in texts mentioning candidates of both genders, followed by texts mentioning only women, and lastly texts mentioning only men. For *Political profiles*, on the other hand, the prevalence of this topic is highest in texts mentioning men and lowest in texts mentioning both men and women, with texts mentioning only women in-between.

The differences are more clearly visible in Figure 5. The topic *Election campaign* is significantly more prevalent when articles mention only women than when they mention only men. By contrast, *Personality* and *Political profiles* are more frequent in texts mentioning only men than in texts mentioning only women. *Energy* and *Immigration* correlate with mentions of female politicians, likely because the two female Federal Councillors, Simonetta Sommaruga and Doris Leuthard, are in charge of these policy areas.

The next step of the analysis looks at how the words associated with each topic correlate with gender. The results are shown in 6 and 7. We refrain from interpreting these results in detail at this stage. We simply note there are no major gender differences in the content of the topics. If this conclusion were confirmed in further analyses, it would mean that, within each topic, male and female candidates are discussed roughly in the same way, or at least in ways that do not reflect clear gender biases.

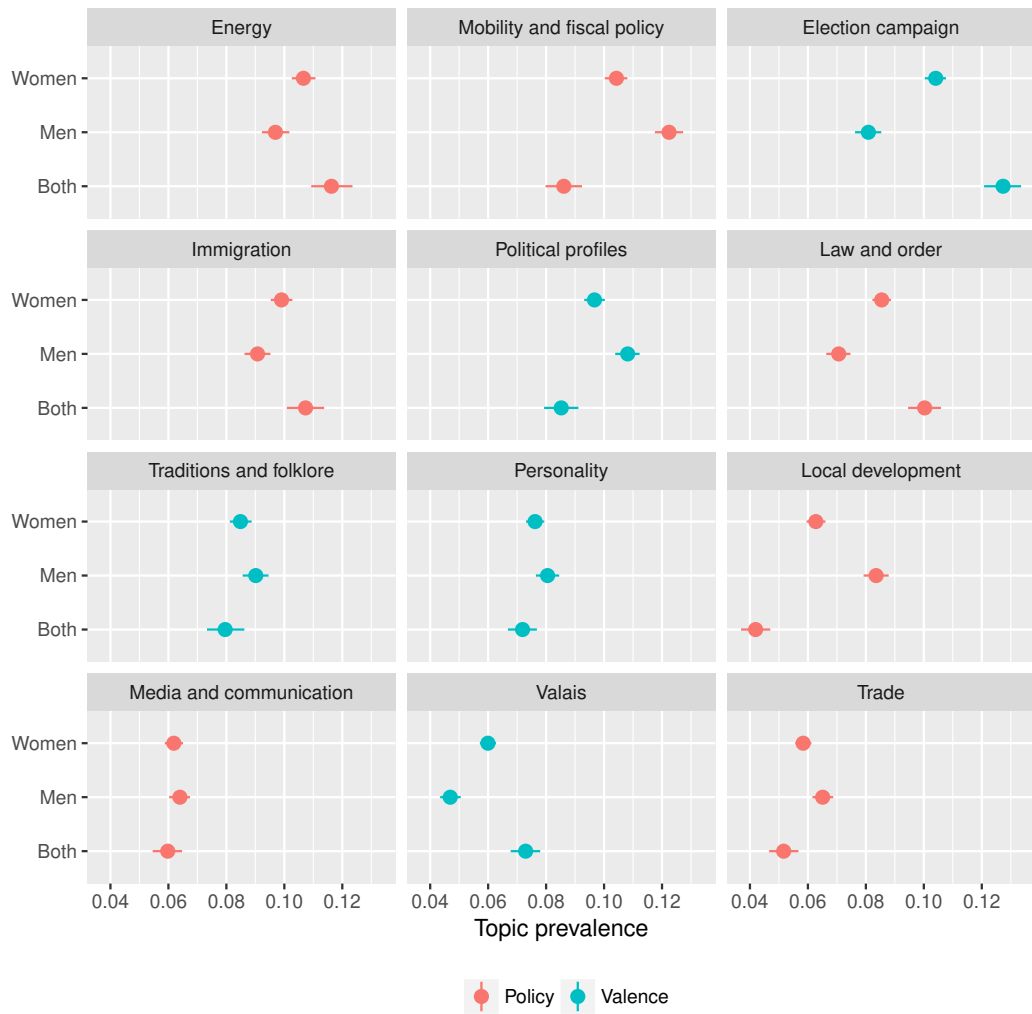


Figure 4: *Topic prevalence and gender (probabilities).*

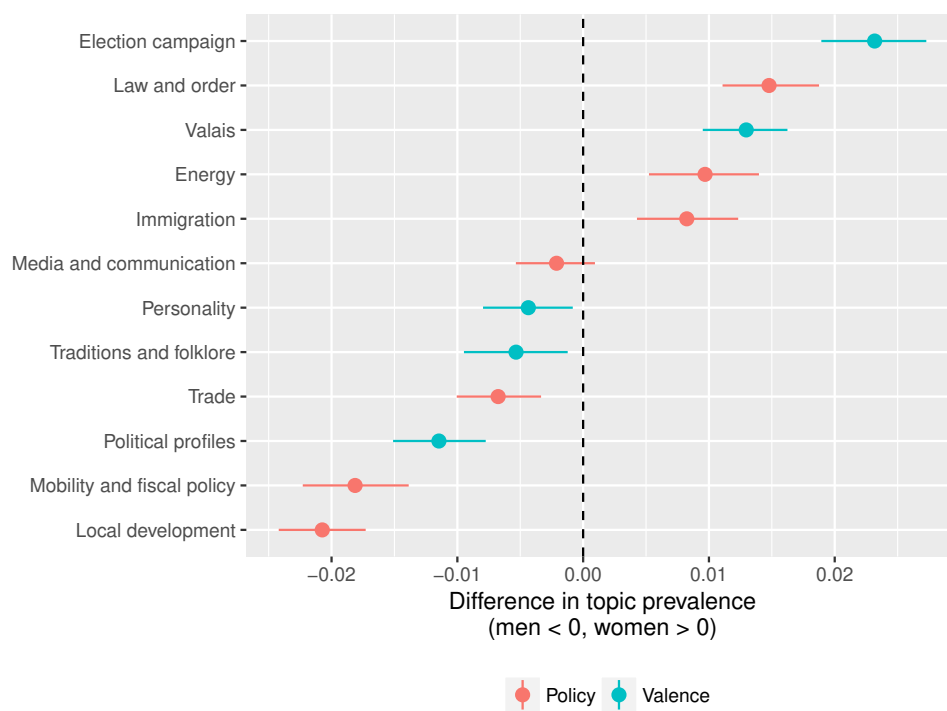


Figure 5: *Topic prevalence and gender (difference in probabilities).*

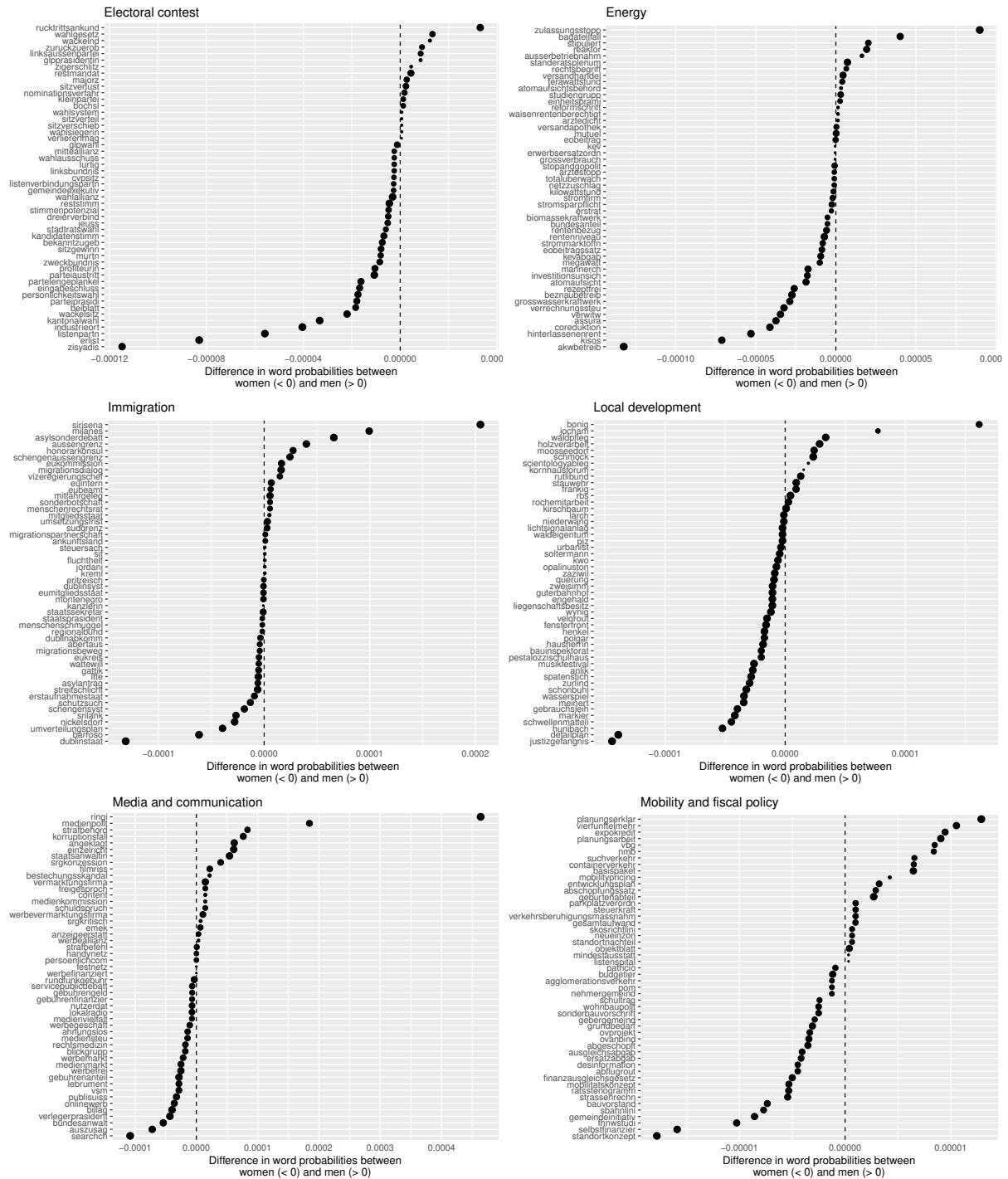


Figure 6: Topic prevalence and gender (difference in probabilities).

5 Conclusion

In this paper, we have shown that media attention is not biased against female candidates as long as they are incumbent. By contrast, male non-incumbents receive more attention than female non-incumbents. It could be however, that this difference is linked with the viability of male and female candidates. In a next version, we will control for candidates' rank on party lists. We expect a strong correlation, but it will be interesting to see whether gender differences will persist once controlling for this important variable.

Second, we have shown that some topics are discussed more frequently in connection with female candidates, while others correlate with mentions of male candidates. Because this is the first output we could produce from our structural topic models, we have refrained from interpreting the results in detail. We will have to make sure that the topics are coherent enough. Moreover, we plan to remove mentions of Federal Councillors, which have likely driven the correlations with some topics. Federal Councillors are not candidates and actually do not belong in this analysis. They are a leftover from another project for which their mentions were important.

Third, it seems that there are few gender differences in the language used for the different topics. Because the analysis is preliminary, we do not wish to put too much emphasis on these specific finding. We plan on expanding this part of the analysis considerably.

References

- Baitinger, Gail. 2015. "Meet the press or meet the men? Examining women's presence in American news media." *Political Research Quarterly* 68(3):579–592.
- Banwart, Mary Christine, Dianne G Bystrom and Terry Robertson. 2003. "From the primary to the general election: A comparative analysis of candidate media coverage in mixed-gender 2000 races for governor and US Senate." *American Behavioral Scientist* 46(5):658–676.
- Blei, David M. and John D. Lafferty. 2007. "A correlated topic model of science." *The Annals of Applied Statistics* pp. 17–35.
- Bystrom, Dianne G, Terry A Robertson and Mary Christine Banwart. 2001. "Framing the fight an analysis of media coverage of female and male candidates in primary races for governor and US senate in 2000." *American Behavioral Scientist* 44(12):1999–2013.
- Druckman, J. N. 2005. "Media Matter: How Newspapers and Television News Cover Campaigns and Influence Voters." *Political Communication* 22(4):463–481.
- Dunaway, Johanna, Regina G Lawrence, Melody Rose and Christopher R Weber. 2013. "Traits versus issues: How female candidates shape coverage of senate and gubernatorial races." *Political Research Quarterly* 66(3):715–726.
- Escobar-Lemmon, Maria C, Valerie Hoekstra, Alice Kang and Miki Caul Kittilson. 2016. "Just the Facts? Media Coverage of Female and Male High Court Appointees in Five Democracies." *Politics & Gender* 12(2):254–274.
- Fiechtner, Stephanie, Philomen Schönhagen and Manuel Puppis. 2016. "Gender und Medien im Vorfeld der eidgenössischen Wahlen 2015." Universität Freiburg.
- Fowler, Linda L and Jennifer L Lawless. 2009. "Looking for sex in all the wrong places: Press coverage and the electoral fortunes of gubernatorial candidates." *Perspectives on Politics* 7(3):519–536.
- Gershon, Sarah. 2012. "When race, gender, and the media intersect: Campaign news coverage of minority congresswomen." *Journal of Women, Politics & Policy* 33(2):105–125.
- Grimmer, Justin and Brandon M Stewart. 2013. "Text as data: The promise and pitfalls of automatic content analysis methods for political texts." *Political Analysis* 21:267–297.
- Hayes, Danny and Jennifer L Lawless. 2015. "A non-gendered lens? Media, voters, and female candidates in contemporary congressional elections." *Perspectives on Politics* 13(1):95–118.
- Hayes, Danny, Jennifer L Lawless and Gail Baitinger. 2014. "Who cares what they wear? Media, gender, and the influence of candidate appearance." *Social Science Quarterly* 95(5):1194–1212.
- Heldman, Caroline, Susan J Carroll and Stephanie Olson. 2005. "'She brought only a skirt': print media coverage of Elizabeth Dole's bid for the Republican presidential nomination." *Political Communication* 22(3):315–335.
- Jalalzai, Farida. 2006. "Women Candidates and the Media: 1992-2000 Elections." *Politics & Policy* 34(3):606–633.
- Kahn, Kim Fridkin. 1994. "The distorted mirror: Press coverage of women candidates for statewide office." *The Journal of Politics* 56(1):154–173.

- Kahn, Kim Fridkin and Edie N Goldenberg. 1991. "Women candidates in the news: An examination of gender differences in US Senate campaign coverage." *Public Opinion Quarterly* 55(2):180–199.
- Kittilson, Miki Caul and Kim Fridkin. 2008. "Gender, candidate portrayals and election campaigns: A comparative perspective." *Politics & Gender* 4(3):371–392.
- Laver, Michael and Ernest Sergenti. 2011. *Party competition: An agent-based model*. Princeton: Princeton University Press.
- Lühiste, Maarja and Susan Banducci. 2016. "Invisible women? Comparing candidates' news coverage in Europe." *Politics & Gender* 12(2):223–253.
- Norris, Pippa. 2000. *A virtuous circle: Political communications in postindustrial societies*. New York: Cambridge University Press.
- O'Callaghan, D., D. Greene, J. Carthy and P. Cunningham. 2015. "An analysis of the coherence of descriptors in topic modeling." *Expert Systems with Applications* 42:5645–5657.
- Roberts, M. E., B. M. Stewart and D. Tingley. 2014. *stm: R Package for Structural Topic Models*.
- Roberts, M. E., B. M. Stewart, D. Tingley, C. Lucas, J. Leder-Luis, S. Gadarian, B. Albertson and D. Rand. 2014. "Structural topic models for open-ended survey responses." *American Journal of Political Science* 58:1064–1082.
- Roberts, Margaret E., Brandon M. Stewart and Edoardo M. Airolidi. 2016. "A model of text for experimentation in the social sciences." *Journal of the American Statistical Association* forthcoming.
- Smith, Kevin B. 1997. "When all's fair: Signs of parity in media coverage of female candidates." *Political Communication* 14(1):71–82.
- Strömbäck, J. and L. L. Kaid. 2008. *The Handbook of Election News Coverage Around the World*. New York, NY: Routledge.
- Ward, Orlanda. 2016. "Seeing Double: Race, Gender, and Coverage of Minority Women's Campaigns for the US House of Representatives." *Politics & Gender* 12(2):317–343.
- Wüest, Bruno, Sarah Bütikofer, Fionn Gantenbeib and Adrian van der Lek. 2016. Der Wahlkampf 2015 in den Schweizer Medien (Selects Medienanalyse 2015). Technical report Institut für Politikwissenschaft, Universität Zürich.

A1 Newspaper corpus

Name	N corpus	N filtered	Type	Name	N corpus	N filtered	Type
20 minuten	6'278	342	freesheet	Neue Luzerner Zeitung	8'933	1'048	national
20 minuten online	5'018	380	news portal	Neue Zürcher Zeitung	8'528	648	national
Aargauer Zeitung	6'922	925	regional	NZZ am Sonntag	1'726	165	sunday
Anzeiger von Uster	288	27	local	Obersee Nachrichten	691	38	local
Basellandschaftliche Zeitung	1'772	306	regional	Oltner Tagblatt	1'173	113	local
Basler Zeitung	8'000	915	national	Ostschweiz am Sonntag	1'045	98	sunday
Basler Zeitung Newsnet	900	127	news portal	Rümlanger	147	16	local
Berner Zeitung	8'758	1'013	national	Schweiz am Sonntag	2'535	373	sunday
Berner Zeitung Newsnet	3'970	494	news portal	Schweizer Bauer	1'713	181	company
Bieler Tagblatt	2'738	165	local	Schweizer Familie	557	11	tabloid
Bilanz online	155	12	business	Schweizer Illustrierte	582	78	tabloid
Blick	4'137	274	boulevard	Seetaler Bote	777	71	local
Blick am Abend	3'102	257	boulevard	Solothurner Zeitung	2'486	356	regional
Bote der Urschweiz	4'793	436	local	Sonntagsblick	1'249	122	sunday
Bündner Tagblatt	4'232	540	regional	Sonntagszeitung	1'232	150	sunday
Cash	10'627	467	business	SRF	7'267	738	news portal
Coopzeitung	569	7	company	St. Galler Tagblatt	15'410	1'302	national
Das Magazin	123	5	weekly	Südostschweiz	5'504	599	national
Der Bund	5'520	664	regional	Swissinfo	153	40	news portal
Der Bund Newsnet	1'527	270	news portal	Tagblatt der Stadt Zürich	324	14	local
Der Landbote	5'010	532	regional	Tages-Anzeiger	6'566	692	national
Die Weltwoche	593	163	weekly	Tages-Anzeiger Newsnet	7'383	760	news portal
Die Wochenzeitung	399	92	weekly	Tageswoche	242	42	regional
Finanz und Wirtschaft (FuW)	1'158	18	business	Tageswoche online	734	105	news portal
FuW online	1'875	21	business	Thurgauer Zeitung	5'088	522	regional
Freiburger Nachrichten	2'818	208	regional	Volketswiler	161	14	local
Furttaler	413	26	local	Walliser Bote	3'899	574	regional
Glattaler	562	70	local	Werdenberg. & Obertoggenb.	5'357	403	local
Glückspost	816	3	tabloid	Willisauer Bote	1'485	142	local
Handelszeitung	773	69	business	Zentral plus	422	104	news portal
Handelszeitung online	503	23	business	Zentralschweiz am Sonntag	978	115	sunday
Infosperber	180	25	news portal	Zofinger Tagblatt	2'272	169	local
Limmattaler Zeitung	1'277	163	local	Zürcher Oberländer	3'934	397	regional
Medienwoche	39	5	news portal	Zürcher Unterländer	3'231	365	regional
Migros-Magazin	897	20	company	Zürichsee-Zeitung	5'081	475	local
Total					205'607	20'104	

A2 Topic model coherence and discrimination

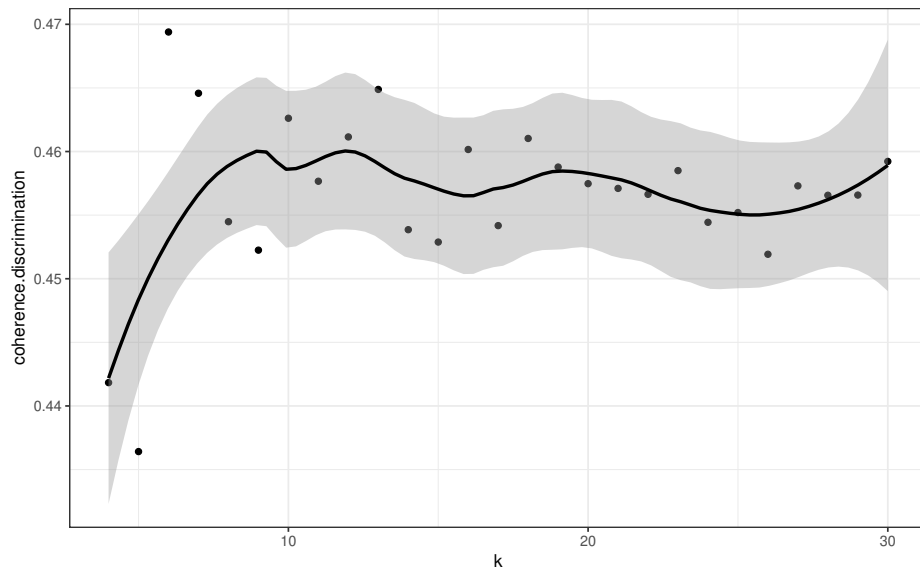


Table A1: *Word2vec* topic coherence and discrimination averages for varying numbers of topics.